

Régiók és alakzatok: Vizuális objektumok kinyerése és illesztése
(From Regions to Shapes: The Extraction and Alignment of Visual Objects)

Válasz Csetverikov Dmitrijnek

az MTA doktori értekezés opponensének

Kató Zoltán

2014. május 16.

Mindenekelőtt köszönöm Prof. Csetverikov Dmitrijnek szakértő, gondos és alapos bírálatát. A bírálat megjegyzéseire és kérdéseire az alábbiakban adok részletes választ.

Kérdésekre, megjegyzésekre adott válaszok

Nem világos, hogy miért csak a paraméteres aktív kontúrok kerülnek említésre, amelyek - a jelölt által javasolt megoldásokkal szemben - valóban nem tudják kezelni a topológiai változásokat.

Több olyan modell is megemlítsékre került, amelyek képesek a topológiai váltás kezelésére (pl. [74], [170], [176]). A jelenleginél részletesebben azért nem tárgyaltam ezeket a módszereket, mert az értekezés fő irányvonala a valószínűségi modellek, és ezen belül az MRF modellek tárgyalása volt.

A másik megjegyzésem az, hogy a javasolt illesztési módszerek működéséhez szükséges pontos binarizálás (szegmentálás) a gyakorlatban legalább olyan nehéz feladat, mint az illesztéshez gyakran felhasznált, stabil jellemzők kinyerése.

Ez függ a konkrét alkalmazástól is - bizonyos típusú képek szegmentálása sokkal egyszerűbb, mint megbízható pontmegfeleltetések előállítása. Erre az értekezés 6.5 ábráján konkrét példát is adok. Másrészt egy konkrét képfeldolgozási folyamatnak általában része a szegmentálás, hiszen egy alkalmazás során valamilyen vizuális információt szeretnénk a képekről kinyerni. Ha viszont a szegmentálás amúgy is rendelkezésre áll, akkor a javasolt regisztrációs módszer eleve hatékonyabb lehet. Ugyanakkor a takarások kezelésében a pont alapú módszerek nyilvánvalóan hatékonyabbak, hiszen ott nincs szükség a teljes alakzatok megfeleltetésére, elegendő, ha megfelelő számú pontpárt találunk. Ezért a kétfajta megközelítés inkább egymás komplementerének tekinthető.

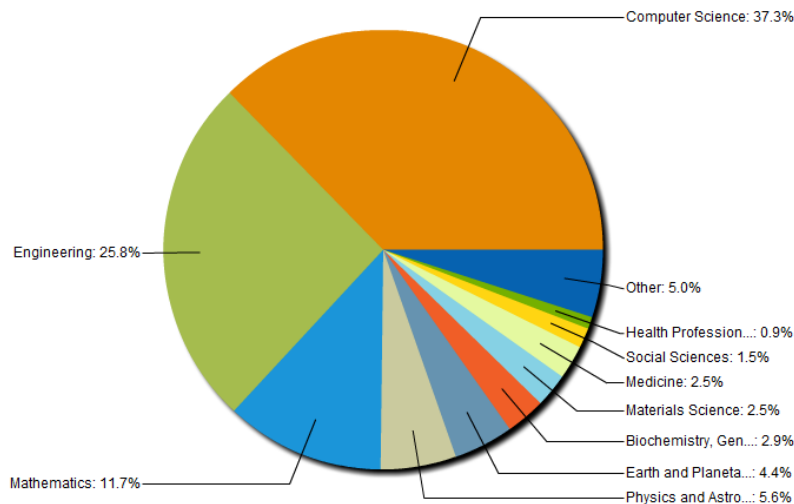
Kritikus megjegyzésként felvehető, hogy kölcsönhatás típusok bővítése ellenére ezekbe a módszerekbe geometriai jellegű információ (pl. az elvárt alakzat jellege)

nehezen építhető be.

Amikor ezek az eredményeim születtek ('90-es évek első fele), ez a terület még közel sem volt ilyen jól kidolgozva, hiszen a Markov modellek csak az 1984-ben megjelent Geman-Geman cikkkel ([97]) kerültek az érdeklődés középpontjába. Köszönhető ez nem kis mértékben a cikkben javasolt hatékony *Gibbs Sampler* energiaminimalizációs eljárásnak. Az akkori hardver számítási kapacitása nehezen bírta volna el az alakzatmodellekkel járó megnövekedett számítási komplexitást, sőt a nagyobb méretű képek kezelése is komoly problémát jelentett. Így az akkori kutatások fő iránya a hatékony energiaminimalizálásra és a kevésbé komplex de hatékony MRF modellek kidolgozására irányult. Ebbe a trendbe illeszkednek az én akkori eredményeim is.

A jelölt által hangsúlyozott alkalmazások és a szoftver oltalom ugyan megtörténtek, de a kapcsolódó hivatkozásokban kevés jel utal arra, hogy a módszerek a gyakorlatban is elterjedtek volna: a hivatkozások döntő része „belterjes”, azaz a saját szakterületekről, nem pedig teljesen más, alkalmazott szakterületekről származik.

Ami a hivatkozásokat illeti, a Scopusban a cikkeim független hivatkozásainak tématerületek szerinti statisztikája nem tűnik „belterjesnek”, sőt abban a *Computer Science* csak 37.3%-al szerepel, s a tágabb *Engineering* terület is mindössze a hivatkozások 25.8%-t adja. A hivatkozások negyede pedig kifejezetten más alkalmazási területekről érkezett.



1. ábra. A cikkeim független hivatkozásainak tématerületek szerinti statisztikája a Scopus-ban

A 38. oldalon például nem világos, mi az F-measure, amely rangsorolja a (kevés) teszteredményt.

A teszteredmények kvantitatív kiértékelése a *Berkeley Segmentation Dataset* (BSDS) alapján készült, melyről bővebb információ található a benchmark honlapján: [http:](http://)

//www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/. A képek száma 100 volt (az adathalmaz vegyesen tartalmaz színes és szürkeárnyaltos képeket, ebből mi csak a színeseket tudtuk felhasználni). Az *F-measure*, amely a BSDS által használt mérőszám az algoritmusok szegmentálási teljesítményének az összehasonlítására, a *precision-recall* görbe alapján egyetlen számmal jellemzi az algoritmus teljesítményét. Maga az *F-measure* a *precision* és *recall* harmonikus középértéke.

A 2.4. ábra sem ad tiszta képet, hogy melyik a jobb eredmény, nem beszélve arról, hogy az ember először felismer, aztán szegmentál, a gép pedig fordítva teszi. (A tesztképeken könnyen felismerhető, de nehezebben szegmentálható objektumok láthatók.) Ráadásul a javasolt módszer nagyon lassú. Hasonlóképpen, a 2.5, 2.6 és 2.10 ábrák sem adnak tiszta képet a teszteredmények minőségéről és általánosíthatóságáról.

A szegmentálási algoritmusok kiértékelése kétségtelenül nehéz kérdés. Az utóbbi évtizedben sok erőfeszítés történt az algoritmusok objektív összehasonlíthatóságára. Ennek egyik eredménye a fentebb már említett *Berkeley Segmentation Dataset*. Ugyanakkor a kvantitatív kiértékelés mellett fontos a tipikus eredmények megjelenítése is. Ezek az ábrák elsősorban illusztrációk, melyek egy-egy tipikus szegmentálási eredményt mutatnak a tesztalmazból.

Ami az algoritmus sebességét illeti, valóban nagyobb számításigénye van az RJMCMC eljárásnak, mint egy hagyományos szimulált hűtésnek. Azonban azt is érdemes figyelembe venni, hogy míg a korábbi algoritmusok igényeltek felhasználói inputot, addig az RJMCMC algoritmus teljesen automatikus, tehát futtatható köteget feldolgozásként is, ahol a futásidő kevésbé kritikus mint egy interaktív rendszerénél.

A módszerek akkor működnek igazán látványosan, amikor nagyon sok objektum jelenik meg egy képen, mint például fakoronák egy légi felvételen vagy sejtek egy orvosi képen. Megjegyzem, hogy ezekre a feladatokra sok hatékony eljárás létezik, különösen az orvosi képfeldolgozásban. Az értekezésben azonban csak egy olyan klasszikus MRF alapú módszerrel történik összehasonlítás, amely az egyik, egyszerűnek tűnő esetben rossz eredményt produkál.

A 3. és 4. fejezetek középpontjában egy eredetileg HOAC-ként megfogalmazott modell MRF ekvivalensének kidolgozása áll. Így a disszertációban elsősorban arra mutattam eredményeket, hogy az így kapott MRF alakzatmodell mennyit javít egy hagyományos, alakzatmodell nélküli módszeren. Ahogyan a bíráló is megemlítette, kevés próbálkozás történt MRF alakzatmodellek használatára. Az említett fejezetek nem csak egy ilyen megoldást mutatnak be, hanem egy általánosabb ekvivalenciát is bizonyos típusú aktív kontúr és MRF modellek között. Megjegyzem, hogy ezzel a kérdéssel már a kezdetekben foglalkoztak a kutatók (ld. Mumford és Shah cikkét [158]), de eddig konkrét modelleket tudtommal még nem dolgoztak ki ezen ekvivalencia kihasználásával. Visszatérve a tesztekkel kapcsolatos bírálói kérdésre: fakoronák szegmentálására léteznek módszerek, néhány ilyen módszerrel történt összehasonlítás a [13] cikkünkben is, de sajnos ezekhez a módszerekhez nehézkes működő implementációt beszerezni, aminek hiányában viszont kvantitatív összehasonlításra nincs

lehetőség. A mikroszkópiai képek esetében a *CellProfiler* egy elterjedt eszköz. Ez azonban sok manuális munkát igényel a felhasználótól (paraméterek finomhangolása, feldolgozási sor összeállítása), míg a mi módszerünk lényegében a kör átmérőn és a Gauss paramétereken kívül nem igényel további paramétereket, automatikusan végzi a képek feldolgozását. A körszerűséggel kapcsolatos általános bírálói megjegyzés szokot még lenni a Hough transzformáció használata. Ezzel az a probléma, hogy nem az alakzatok valódi határvonalát adja, hanem egy tökéletes körként detektálja azokat. Ezzel meg lehet határozni, hogy hány alakzat található egy képen, de ha azok pontos területére vagy más alakjellemzőjére vagyunk kíváncsiak (fakoronák esetében ez pl. nagyon fontos), akkor már egy ilyen eredmény nem használható a feladat megoldására.

Ehhez azonban a többretegű modell esetében az eljárást lényegesen fel kell gyorsítani, hiszen jelenleg egy kisméretű képre akár 1000 mp lehet a futási idő, ami a gyakorlatban elfogadhatatlan.

Ahogy a bíráló is megjegyezte, a többretegű modell még egy friss eredmény, így az algoritmus optimalizálására eddig kevés energiát fordítottunk. A kérdés napirenden van, és elsősorban egy hatékonyabb, nem pont-alapú mintavételező eljárás kidolgozásában látjuk a gyorsítás lehetőségét. Jelenleg ebben az irányban folynak a kutatások.

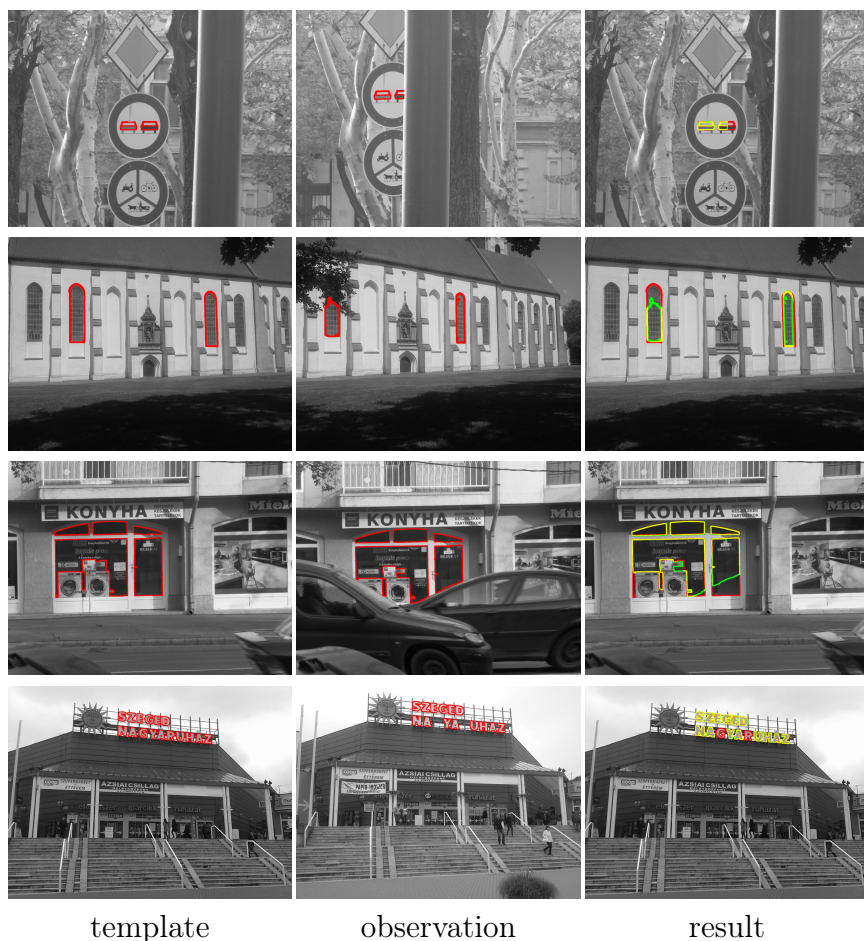
Lehet, hogy nem olvastam elég figyelmesen a szöveget, de nem találtam magyarázatot arra, hogy a gyakorlatban fontos térbeli alakzatok itt miért maradtak ki. Feltehetően azért, mert több szabadságfok esetében nagyobb a valószínűsége, hogy egy nemlineáris módszer beragad egy lokális minimumban és hamis eredményt ad.

A 6. fejezetben tárgyalt módszer kiterjesztése 3D alakzatokra időközben megtörtént. Ez részletesen azért nem került be a dolgozatomba, mert csak 2011-ben kezdtünk el a témán dolgozni, és az értekezés írásakor még nem álltak rendelkezésre kellően publikált eredmények. A 6.6.5. fejezetben azonban röviden írok erről a kiterjesztésről, és néhány eredmény is látható a 6.11 ábrán tüdő CT képek illesztésére. Azóta a módszert alaposan teszteltük és hatékony numerikus algoritmust is adtunk az egyenletrendszer megoldására. Ezeket az eredményeinket Santa Zsolt hallgatómmal közös cikkben publikáltuk a 2013 júniusi CVPR konferencián ([48] hivatkozás). A módszer stabil eredményeket ad és a több száz teszteset kiértékelése során nem tapasztaltuk, hogy lokális minimumban akadna el.

A megfeleltetés nélkül, integrális jellegű módszerek közismert hátránya azonban a hiányos adatokkal szembeni kisebb robusztusság, különösen a magasrendű nyomatékokat alkalmazó módszerek esetén.

Az integrálás miatt a takarások kezelése valóban problémás a javasolt regisztrációs módszer számára - ezt többször is megemlítem a robusztusság kiértékelésénél. Ugyanakkor erre a problémára is születtek azóta új eredményeink. Konkrétan lineáris transzformációk esetében van megoldásunk a takarások kezelésére, melyet éppen

most fogadtak el az ICPR 2014 konferenciára (Zsolt Santa and Zoltan Kato. *Affine Alignment of Occluded Shapes. In Proceedings of International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, August 2014. IEEE. (Accepted)*). A cikkben javasolt módszer a szintetikus adathalmazon végzett kiértékelések alapján 20% mértékű takarásokat még jól kezel. Az alábbi ábra is ebből a cikkből való.



2. ábra. Results on real images. The first two columns are containing the *template* and the *observation* images, respectively with the contours of the segmented shapes. In the last column the results are visualized, where the contour of the *template* shape is denoted by red, the contour of the transformed *observation* is denoted by green and the intersection is denoted by yellow colours.

A téziseket elfogadom azzal a megjegyzéssel, hogy az 1.2, 2.3. és különösen a 2.4. tézist kevés publikáció támasztja alá, ezért ezen tézisek a súlya kisebb. A 2.3. tételhez kapcsolódó fő publikációban a jelölt a négy társszerző közül a harmadik, ez tovább csökkenti a tétel súlyát.

A tézisekhez kapcsolódó publikációk száma valóban változatos. Néhol a konferenciatickek kínáltak megfelelő fórumot, máshol viszont egy nagyobb folyóiratcikk tűnt a legjobb publikációs lehetőségnek.

Az 1.2 tétel a többhőmérsékletes szimulált hűtés algoritmusát és annak elméleti bizonyítását tartalmazza, ezért ez inkább elméleti jellegű eredmény. Az eredménye-

ket főként a [24] folyóiratcikkben publikáltam, amire viszont a mai napig rendszeresen kapok hivatkozásokat és ez az egyik legtöbbet hivatkozott cikkem. A kapcsolódó publikációk száma valóban kevesebb, mint a többi tézisnél, de a hivatkozások száma talán jelzi, hogy az ebben megfogalmazott eredmények fontosak és a mai napig hasznosak.

A 2.3 és 2.4 tézisek még új kutatási területek, igazán aktívan 2009 óta foglalkozok a modellek MRF ekvivalenciájával és az így megkonstruált algoritmusok kidolgozásával. Ezért az eredmények publikáltsága valóban kisebb, mint a régebbi, vagy a párhuzamosan de aktívabban futó regisztrációs téma publikáltsága. Megjegyzem, hogy a 3. fejezetben tárgyalt MRF modell és a kapcsolódó levezetések egy most készülő folyóiratcikk fő tartalmát adják.

Szeged, 2014. május 16.

Kató Zoltán